

# 基于知识进化算法的读者满意度评价<sup>①</sup>

吴凤娟

(商丘师范学院 计算机科学系, 商丘 476000)

**摘要:** 通过结合知识进化论与生物进化论思想, 提出了知识进化算法, 给出该算法的基本原理和实现途径, 并创建了知识的评价函数。把该算法用于图书馆读者满意度评价实例中, 可获得成功的试验结果。这表明知识进化算法可促进知识的量与质的提升, 为正确决策提供有效的辅助手段。

**关键词:** 知识进化论; 传承算子; 创新算子; 读者满意度

## Evaluation of the Reader's Satisfaction Based on the Knowledge Evolution Algorithm

WU Feng-Juan

(Computer Science Department, Shangqiu Normal College, Shangqiu 476000, China)

**Abstract:** Trough combining evolutionary epistemology theory with biological evolution theory, this paper presents a knowledge evolution algorithm. The basic principle and realizing ways is also given. The evaluating function of the knowledge is created too. When the algorithm is used in the evaluation of the reader's satisfaction in a library, we can obtain successful experiment results. The experiment shows that the knowledge evolution algorithm can promote both the quantity and the quality of the knowledge, which is an effective assistant means to make a correct decision.

**Key words:** knowledge evolutionary epistemology; inheriting operator; innovation operator; reader's satisfaction

### 1 引言

科学发展与人类进步都沿着两条类似的进化路线进行<sup>[1]</sup>, 即自然进化和知识进化, 这是促使科学知识能更快更新和增长的机制。著名的科学哲学大师卡尔·波普尔提出的知识进化论<sup>[2]</sup>认为知识进化与生物进化一样, 也存在群体与个体进化, 个体进化中突变的累积效应最终导致和构成了群体进化的结果。

受知识进化论的启发, 通过借鉴生物进化论思想, 提出了一种解决知识优化的新方法——知识进化算法, 在图书馆读者满意度评价中的应用结果显示, 该算法能有效模拟知识进化机制, 对获取新知识实效显著。

### 2 知识进化算法原理及其实现

知识进化算法借鉴“适者生存”的原理, 对当前知识进行传承与创新, 使知识得以进化, 并由知识的评

价约束了进化的方向。主要包含知识的表示及编码、传承算子、创新算子以及知识的评价。

#### 2.1 知识的表示及编码方法

##### 2.1.1 知识表示

知识表示把自然语言描述的知识转变成计算机可利用的形式。知识表示常用方法有谓词逻辑、产生式规则、语义网络、框架、神经网络等。因产生式规则简单清晰、易于理解, 本文选其作为知识表示形式, 一般形式为: IF  $E_1 \wedge E_2 \dots \wedge E_n$  Then H(结论) CF(R) ( $0 \leq CF \leq 1$ )

其中  $E_i$  为属性, 通过 And( $\wedge$ )或 Or( $\vee$ )连接起来。CF(Certainty Factor)为规则的可信度, 表示规则的不确定强度。产生式规则可由动态知识树<sup>[3]</sup>(Knowledge Tree)表示, 结点由规则的前提属性、结论和操作符组成。根结点是 IF 结点, 其左子树由 And 和前提属性构成, 右子树为结论。如下规则可由图 1 所示知识树表

<sup>①</sup> 基金项目:商丘师范学院青年科研基金(2010QN11)

收稿时间:2011-03-05;收到修改稿时间:2011-04-19

示:

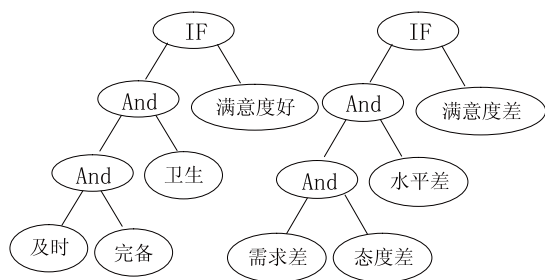


图 1 规则 1 和规则 2 的知识树

规则 1: IF 信息及时 And 信息完备 And 环境卫生 Then 读者满意度好

规则 2: IF 信息需求差 And 服务态度差 And 专业水平差 Then 读者满意度差

### 2.1.2 知识编码方法

由于知识规则间存在较大差异(如条件个数及结论情况不同等),因此知识树在形状、深度及结点数等方面差异较大。采用二进制编码<sup>[4]</sup>(结构如图 2)实现知识树的编码可灵活运用知识规则。图 2 中布尔标志位“标号”表示知识树是否包含该属性或结论,即当标号为 1 时,表示此属性或者结论被包含在知识树中;否则该知识树不包含此属性或者结论。“结论”为知识树中右孩子的编码。“属性值”表示知识树中左孩子属性值的编码。

标号	结论	标号	属性值	标号	属性值	.....
----	----	----	-----	----	-----	-------

图 2 知识规则个体编码结构

对属性值进行二进制编码时所使用到的字符串位数  $n$  符合:

$$n = \lceil \log_2 M \rceil \quad (1)$$

其中,  $M$  表示某一个属性值域内属性值的数量。

### 2.2 知识传承算子

知识的传承就是传递较优知识, 接续较优知识。传承算子的操作步骤为:

- (1) 根据知识评价函数计算知识树  $x_i$  适应度值  $f(x_i)$ ;
- (2) 确定知识规则集中知识的传承规模  $C$ ;
- (3) 在当前知识规则集中按某种策略选出  $C$  条较优的知识规则进入下一代知识规则集中。

如下传承算子伪代码实现把知识个体排序并选择前  $C$  条最优知识规则:

```

individual Inhe_Oper(individual Population[],int C)
//返回传承规模为 C 的知识种群
{for(i=0;i<popsize;i++) { 计算知识规则库
Kno_Rule[]中知识个体  $x_i$  的适应度  $f(x_i)$ ;}
for(i=0;i<popsize;i++) { 知识个体按照适应度值
 $f(x_i)$ 从大到小排列;}
for(i=0;i<C;i++) {知识规则库 Kno_Rule[]中前 C
条知识赋给知识种群 Population[];}
return Population;}
    
```

### 2.3 知识创新算子

知识在进化过程中不断创新, 逐步提升本身价值。创新过程中, 不同知识的交叉、融合是相当频繁的, 同时知识的创新是没有终极目标的。创新算子可产生新的知识规则, 使得求解某类问题的知识得以更新, 操作步骤为:

- ① 对由传承算子得到的知识规则集中的知识个体进行随机配对;
- ② 知识个体对应属性进行重新组合和局部变异;
- ③ 知识个体的有效性判断, 包括知识个体的结构性判断、属性值可行域的判断等。

创新算子的伪代码描述为:

```

individual Inno_Oper(individual Population[C])
{for(i=0;i<C;i++) //初始化知识规则 ID
{index[i]=i;}
for(i=0;i<C;i++) //负责随机配对
{生成比 C 小的随机数 point;
交换 index[i]和 index[point+i]的值;}
for(i=0;i<C-1;i+=2) //负责知识对的基因位重新组
合
{随机生成概率 P;
if(P<Pc){生成比知识规则编码长度小的随机数
point;
for(j=point;j<knowledgelength;j++)
{交换 Population[index[i]]和 Population[index[i+1]]
的部分知识基因;}}}
for(i=0;i<C;i++) //负责知识个体的局部基因突变
{for(j=0;j<knowledgelength;j++)
{随机生成概率 P;
if(P<Pm) {知识个体发生局部基因变异;}}}
for(i=0;i<C;i++) //知识个体的有效性判断
{if(结论标识全为 0) {删除此知识规则;}
    
```

```

else if(属性标识全为 0) {删除此知识规则;}
if(属性值不再属性值域内) {该属性标识置为
0;}}
    
```

在创新算子的设计过程中，知识个体的交叉概率  $P_c$  由知识个体的适应度函数得到，计算公式为：

$$P_c = \begin{cases} 0.9 - k_1 \cdot (f' - f_{avg}), & f' \geq f_{avg} \\ k_2 & f' < f_{avg} \end{cases} \quad (2)$$

式中， $f_{max}$  为群体中最大的适应度值； $f_{avg}$  为每代群体的平均适应度值； $f'$  为要交叉的两个个体中较大的适应度值。系数  $k_1$ 、 $k_2$  取(0,0.9)区间的值。

知识个体的局部变异是产生当前最优知识的重要原因，其主要目的是改变知识树的大小及树中结点所含的信息。变异操作以一定的变异概率  $p_m$  发生， $p_m$  的值由知识个体的适应度函数得到，计算公式为：

$$P_m(x_i) = \begin{cases} 0.1 - \frac{k_3 \cdot (f(x_i) - f_{avg})}{f_{max} - f_{avg}}, & f(x_i) \geq f_{avg} \\ k_4 & f(x_i) < f_{avg} \end{cases} \quad (3)$$

这里， $p_m(x_i)$  为知识个体  $x_i$  的自适应变异率， $f_{max}$  为个体中最大的适应度值， $f_{avg}$  为每代群体的平均适应度值， $f(x_i)$  为要变异个体  $x_i$  的适应度值，系数  $k_3$ 、 $k_4$  取(0,0.1]区间的值。

### 2.4 知识适应度评价方法

采用线性组合法来衡量知识的优劣程度，适应度表达式为：

$$f(x_j) = w_1 \cdot A_j + w_2 \cdot C_j + w_3 \cdot D_j \quad (4)$$

其中， $w_1$ 、 $w_2$  和  $w_3$  是权重，且  $w_1 + w_2 + w_3 = 1$ 。当表达式的参数对解决问题的贡献较大时，需要对其赋予较大的权值，否则，给次要的参数赋予较小的权值。 $A_j$ 、 $C_j$  和  $D_j$  分别表示知识规则  $j$  的正确度、可信度和简洁度<sup>[5]</sup>，定义如下：

(1) 正确度计算公式为：

$$A(r_j) = \frac{k_j}{n} \quad (5)$$

其中， $k_j$  表示实例库中与知识规则  $r_j$  相匹配的实例数目； $n$  表示实例库中实例总数。

(2) 可信度用来表示知识规则的精确程度，计算公式为：

$$C(r_j) = \frac{TP}{TP + FP} \times \frac{TP}{TP + FN} \quad (6)$$

其中， $TP$  表示实例库中与知识规则  $r_j$  的前提条件和结论相匹配的实例数目； $FP$  表示实例库中与知识规则

$r_j$  的前提条件相匹配而与结论不匹配的实例数目； $FN$  表示实例库中与知识规则  $r_j$  的前提条件不匹配而与结论相匹配的实例数目。

(3) 简洁度：一般情况下，知识树越小规则越简洁，预测能力越强。计算公式为：

$$D(r_j) = \frac{\log_2(1 + M_j)}{\log_2(1 + N_j)} \quad (7)$$

其中， $M_j$  表示知识规则  $r_j$  结论的个数； $N_j$  表示知识树叶结点的总数量。

### 2.5 算法的实现步骤及其流程

知识进化算法的实现步骤描述为(流程图如图 3)：

- S1 通过对某个领域实际问题的分析求解，获取知识，并将其与先验知识并入预知识库中；
- S2 对预知识库利用实例库去粗取精，形成知识规则库；
- S3 依据知识的适应度评价函数计算知识规则的适应度；
- S4 使用传承算子与创新算子使知识得以进化，并对得到的新知识规则进行有效性判断；
- S5 判断是否有新的知识规则，如果有，转向 S6；没有，转向 S7；
- S6 计算新知识规则的适应度，将其加入知识规则库中；
- S7 算法结束。

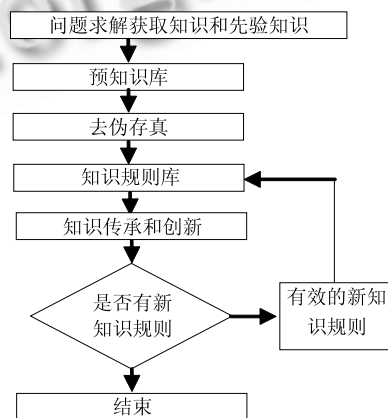


图 3 知识进化算法流程

## 3 实验与分析

以图书馆读者满意度评价中的知识规则进化过程为例来检验知识进化算法的有效性。读者满意度<sup>[6]</sup>指

读者对图书馆所提供的文献信息产品和无形服务的内心感受和主观评价。影响因素主要包括文献、员工、环境三大方面<sup>[7]</sup>。将这些因素作为评价的指标体系,包括表 1 中文献内容质量等 12 项指标,各指标(前提属性)的取值域为{好(good),较好(cgood),一般(normal),差(poor)}。满意度被划分为满意(sat)、比较

满意(csat)、基本满意(bsat)、不满意(usat)4 个层次。前提条件和结论的各个属性的取值相对应的编码均为 00、10、01、11。

通过对获得的图书馆读者满意度评价知识及已有的先验知识进行分析、归纳与检验,形成知识规则库,如表 1 所示。表中,“-”表示无此评价。

表 1 读者对图书馆满意度的知识规则库

规则号	文献内容质量	文献检索难易	现有复本量	馆际互借和文献传递及时性	文献借阅时间	员工服务态度	员工专业水平	员工服务深度	环境布置	环境卫生	环境秩序	读者满意度
1	good	good	good	good	good	cgood	normal	normal	good	normal	cgood	sat
2	cgood	cgood	good	normal	normal	poor	normal	poor	good	good	good	bsat
3	cgood	good	cgood	good	-	normal	normal	-	-	normal	poor	bsat
4	normal	-	good	normal	-	normal	-	good	poor	-	-	usat
5	cgood	good	-	-	good	-	good	-	cgood	poor	good	bsat
6	-	good	-	good	poor	poor	normal	-	-	-	-	usat
7	-	-	good	cgood	-	good	-	good	cgood	-	cgood	csat
8	cgood	-	-	cgood	good	normal	-	-	-	poor	-	bsat
9	cgood	good	cgood	normal	-	-	cgood	-	normal	good	poor	bsat
10	cgood	cgood	good	-	cgood	normal	-	good	cgood	poor	-	sat
11	normal	normal	-	cgood	-	-	normal	-	normal	-	good	csat
12	cgood	-	good	-	good	good	-	normal	normal	-	-	sat

按照图 2 中知识规则的编码结构,对表 1 中的知识规则进行二进制编码并计算出适应度值如表 2 所示。

表 2 知识规则编码及其适应度评价

规则号	知识规则的二进制编码	知识评价
1	100100100100100110101101100101110	0.84
2	10111011010010110111110111100100100	0.82
3	101110100110100000101101000000101111	0.82
4	111101000100101000101000100111000000	0.92
5	10111010000000100000100000110111100	0.86
6	111000100000100111111010000000000000	0.92
7	110000000100110000100000100110000110	0.87
8	101110000000110100101000000000111000	0.90
9	101110100110101000000110000101100111	0.77
10	100110110100000110101000100110111000	0.80
11	110101101000110000000101000101000100	0.88
12	100110000100000100100000101101000000	0.90

设置传承算子中的知识传承规模  $C=6$ , 创新算子

中  $k_1=k_2=0.6$ ,  $k_3=k_4=0.05$ , 正确度、可信度和简洁度的权重分别为 0.5、0.3、0.2。选择适应度高的编号为 4、6、7、8、11 和 12 的 6 个知识个体形成初始知识种群,对其进行一次进化后得到的结果为例,新知识个体编码及其含义如表 3 和表 4 所示。经检验,表 4 中规则 16 是无效规则。将新知识规则 13、14、15、17 和 18 加入到知识规则库中,使知识规则库得以更新,实现知识的进化。经实验验证,该算法满足知识进化的要求,符合预定的知识增长模式。

表 3 新知识规则编码及其适应度评价

规则号	知识规则的二进制编码	知识评价
13	11110100010010110010100000000111100	0.94
14	101110000000110000101010100111000000	0.90
15	111000100100100110000101000101000100	0.45
16	010101101000110001111101000000000000	0
17	110100000100110100100000101101000000	0.83
18	10011000010000000100000100110000010	0.93

表 4 新知识个体

规则号	文献内容质量	文献检索难易	现有复本量	馆际互借和文献传递及时性	文献借阅时间	员工服务态度	员工专业水平	员工服务深度	环境布置	环境卫生	环境秩序	读者满意度
13	normal	-	good	normal	good	normal	-	-	-	poor	good	usat
14	cgood	-	-	cgood	-	normal	-	good	poor	-	-	bsat
15	-	good	good	good	cgood	-	normal	-	normal	-	good	usat
16	normal	normal	-	cgood	-	poor	normal	-	-	-	-	-
17	good	-	good	cgood	good	good	-	normal	normal	-	-	csat
18	cgood	-	good	-	-	good	-	good	cgood	-	-	sat

#### 4 结论

本文将知识进化论与生物进化论思想结合起来,提出的知识进化算法有利于促进知识的进化,符合人类科学知识增长的特点。图书馆读者满意度评价的知识规则进化过程获得了成功的试验结果,表明了知识进化算法的可行性和有效性。今后的主要工作是对该算法的性能进行评估,并拓宽其应用领域。

#### 参考文献

- 1 YongQin Tao, DuWu Cui, TaiShan Yan. Knowledge Evolutionary Algorithm Based on Granular Computing. IEEE Conference on Cybernetics and Intelligent Systems, 2008, 7:1230-1235.

- 2 刘纯青,杨莘元,张颖.知识进化策略.系统工程与电子技术, 2007,29(6):1017-1020.
- 3 刘明吉,王秀峰,王治宝,黄亚楼.一种基于遗传算法的知识挖掘算法.计算机工程,2000,26(8):13-14.
- 4 Wang CH, Hong TP, Tseng SS, et al. Automatically Integrating Multiple Rule Sets in a Distributed-Knowledge Environment. IEEE Trans. on Systems, Man and Cybernetics. 1998,28(3):471-476.
- 5 Ghosh A, Nath B. Multi-objective rule mining using genetic algorithms. Information Science, 2004,163: 123-133.
- 6 臧鸿妹.国外图书馆信息服务质量评价模型研究.情报杂志, 2007,(8):40-42.
- 7 蒋知义,邹凯.基于模糊综合评价的图书馆读者满意度分析.情报学报,2007,26(3):464-469.

(上接第 263 页)

与生产成本会产生重大影响。采购管理的优劣成为企业提高竞争力的关键环节。本文以实施企业精益采购为目标,围绕精益采购信息技术支撑系统开展研究,为开展精益采购提供必要的技术手段。文章分析了航空制造企业供应采购管理现状,提出了符合现代航空制造特点的数字化采购管理模式,设计了数字化采购管理系统系统,涵盖了采购计划、合同管理、合同追踪以及供应商目录管理等。

#### 参考文献

- 1 陈原.民用航空制造业供应链协调管理研究[博士学位论文].长沙:中南大学,2007.
- 2 綦方中,钟凌燕,潘晓弘.敏捷供应链绩效评价过程与方法研究.计算机集成制造系统,2006,12(9):1472-1476.

- 3 尹昌艳.GL 汽车玻璃制造公司采购管理模式改进研究[硕士学位论文].长沙:湖南大学,2007.
- 4 宋利康,李云峰,於建华,郑和兴.基于松耦合架构的航空制造企业集成物流信息链技术研究.航空计算技术,2007,4.
- 5 胡冬伟,刘聆哲,王佳,周跃进.中小离散制造企业采购管理系统研究.工业控制计算机,2007,20(12):62-64.
- 6 王勇.ISC 下汽车制造企业的精益供应研究[硕士学位论文].武汉:武汉科技大学,2007.
- 7 宋红梅,冯志勇,李杰.基于本体的汽车零部件采购订单的自动生成.计算机工程与应用,2007,43(11):228-231.
- 8 刘忠梅.我国船舶制造业供应商管理研究[硕士学位论文].哈尔滨:哈尔滨工程大学,2007.
- 9 陈鸿宇,胡涛,姚路.基于 Agent 的装备采购供应商仿真模型研究.海军工程大学学报,2008,(2):93-97.